Year: 2006

Positional accuracy of biological research data in GIS – A case study in the Swiss National Park

Imfeld, Stephan ; Haller, Ruedi ; Laube, Patrick

Abstract: Original field research data requires information about the positional accuracy of objects located in the field, expecially when analysed in the context of GIS. We present the results of a case study assessing the spatial accuracy of vegetation sampling data. The positional accuracy of a research grid consisting of adjacent squares of 20mx20m set up using a large scale orthophoto (1:2000) used for vegetation studies was assessed using surveying techniques. To study the absolute positional accuracy of the setup, the exact locations of a large part of these squares were determined using surveying techniques. The mean positional error was 5.2m (span 0.9-9.1m) for pegs located in the corners of the squares. The size of the individual squares ranged from 64% to 133% of the planned size of 400m2. The average (horizontal) distance of the true locations (n=335) was exactly as planned (20.0±2.1m). The minimum distance was 14.8m and the maximum distance was 25.4m. The mean horizontal angle in the corners of the plots was 89.9±3.3° (span 77.4-102.3°) (n=615). Overall, 67.4% of the whole area was in accordance to the GIS database, 32.6% was falsely attributed to wrong sampling squares. The influence on vegetation classification statistics was small (maximum of 0.58%). Even with the aid of relatively sophisticated instruments such as orthophotos, the positional accuracy in the original study was low, resulting in differences in plot area of over 200%. Nevertheless, the influence on the results of a single study are moderate. By contrast, these errors are of high concern in areas of intense interdisciplinary research such as national parks. It is thus recommended that for a focus research area as the site under investigation, surveying techniques should be implemented to enable long-term research and to minimize the risk of incorrect research results due to inaccurate spatial data.

Originally published at:
Imfeld, Stephan; Haller, Ruedi; Laube, Patrick (2006). Positional accuracy of biological research data in GIS – A case study in the Swiss National Park. In: 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Lissabon (P), 5 July 2006 - 7 July 2006. ISARA, 275-280.

Positional Accuracy of Biological Research Data in GIS – A Case Study in the Swiss National Park

Stephan Imfeld[1], Ruedi Haller[2] and Patrick Laube[3]

1 Department of Geography, University of Zurich
Winterthurerstr.190, CH-8057 Zurich, Switzerland
Tel.: + 41 44 635 5253; Fax: + 41 44 635 6848
imfeld@geo.unizh.ch

2 Swiss National Park
Chasa dal Parc, CH-7530 Zernez, Switzerland
Tel.: + 41 81 856 1282; Fax: + 41 81 856 1740
rhaller@nationalpark.ch

3 School of Geography and Environmental Science
University of Auckland
Private Bag 92019
Auckland, New Zealand
Tel: +64 9 3737599; Fax: +64 9 3737434
p.laube@auckland.ac.nz

Abstract

Original field research data requires information about the positional accuracy of objects located in the field, expecially when analysed in the context of GIS. We present the results of a case study assessing the spatial accuracy of vegetation sampling data. The positional accuracy of a research grid consisting of adjacent squares of 20mx20m set up using a large scale orthophoto (1:2000) used for vegetation studies was assessed using surveying techniques. To study the absolute positional accuracy of the setup, the exact locations of a large part of these squares were determined using surveying techniques. The mean positional error was 5.2m (span 0.9-9.1m) for pegs located in the corners of the squares. The size of the individual squares ranged from 64% to 133% of the planned size of 400m2. The average (horizontal) distance of the true locations (n=335) was exactly as planned (20.0±2.1m). The minimum distance was 14.8m and the maximum distance was 25.4m. The mean horizontal angle in the corners of the plots was 89.9±3.3° (span 77.4-102.3°) (n=615). Overall, 67.4% of the whole area was in accordance to the GIS database, 32.6% was falsely attributed to wrong sampling squares. The influence on vegetation classification statistics was small (maximum of 0.58%). Even with the aid of relatively sophisticated instruments such as orthophotos, the positional accuracy in the original study was low, resulting in differences in plot area of over 200%. Nevertheless, the influence on the results of a single study are moderate. By contrast, these errors are of high concern in areas of intense interdisciplinary research such as national parks. It is thus recommended that for a focus research area as the site under investigation, surveying techniques should be implemented to enable long-term research and to minimize the risk of incorrect research results due to inaccurate spatial data.

Keywords: spatial accuracy, error analysis, uncertainty, environmental data, model

1 Introduction

The Swiss National Park (SNP) was founded in 1914 in the eastern alpine parts of Switzerland. In 1916 the importance of long-term research was already realized (WNPK, 1916). Since then a large amount of research has been conducted in the SNP. In 1992 the development of a database using a geographical information system was started with the aim to serve as an integrating platform for monitoring, research and administrative purposes (Allgöwer and Bitter, 1992). As this database and its use are growing, questions about the positional accuracy become important. Multiple and especially interdisciplinary use of original research data requires information about the positional accuracy of objects located in the field (Rettie, 1999; Lexer et al., 2004).

The work presented here is an assessment of the positional accuracy of field data originating from a biological study conducted on the Alp Stabelchod (Achermann, 2000; Schütz et al., 2003). The site is an area of sub-alpine grassland of about 10ha located at an altitude of 1950m a.s.l. (range 1920-1980m a.s.l.). As this area is a research focus in the SNP (e.g. Braun-Blanquet, 1931; Krüsi et al., 1995; Holzgang, 1997; Bigler, 1997; Schütz et al., 1998; Besson, 1999; Leuzinger, 1999; Maggini, 1999; Achermann, 2000), a precise spatial database has to be established. As the SNP and especially the Alp Stabelchod is visited by a lot of tourists, permanent installations of pegs and poles to identify hundreds of research plots are to be avoided. Nevertheless it is a requirement for longtime research to be able to find the exact location of previous data collections. Additionally the accuracy of area calculations can be improved as will be shown in this study. As a first action the spatial accuracy of the above mentioned study (Achermann, 2000) is evaluated. Consecutively it will be the aim to determine the exact locations of all research plots that are still reconstructable.

Accuracy of spatial data has several different aspects (semantic, thematic and classification, measurement, modeling etc.) for each of which a separate body of literature can be found (e.g. Goodchild, 1995; Salgé, 1995; Goodchild and Gopal, 1989). The focus of interest here is the positional accuracy and the measurements based on the locations. Several methods have been developed to assess it for point objects, lines and polygonal structures (e.g. Goodchild and Hunter, 1997; Skidmore and Turner, 1992; Sbresny, 1997). In this paper we are mainly dealing with point objects, for which measurements of error are straight forward. The handling of error as described in Hunter and Goodchild (1995) and Goodchild and Gopal (1989) may not be so important here in the future, because for such a small area as the Alp Stabelchod (10ha) it should be no problem to achieve positional accuracy of a few centimeters by using surveying techniques.

2 Methods

In 1997 Achermann (2000) investigated the influence of red deer upon the subalpine grassland ecosystem of the Alp Stabelchod in the Swiss National Park. Achermann's study used a grid set up across the whole area of approximately 10 hectares. An orthophoto (scale 1:2'000) was produced on the basis of an infrared aerial photo (1:7'000). With the aid of this orthophoto a grid of 20m x 20m plots was set up in the field. The grid was oriented according to the Swiss coordinate system. The corners of the plots were located in the field and marked with wooden pegs. In this manner 268 plots were set up in 1997 using a total of 280 pegs. The plots were used for sampling of vegetation parameters, soil chemistry and habitat use assessment of red deer in a spatially defined manner.

In July 2000 the exact location of a large part of these pegs was determined by the authors using a surveying instrument (Leica TCA1102). The coordinates of the reference points were determined in 1998 with a surveying GPS with an horizontal accuracy of 7mm. The positional accuracy of the measurements of the pegs is within a few centimeters. With this data, the location of 194 pegs, the length of 335 distances between pegs and the size of 141 plots could be determined and compared to the planned set up.

The geometrical analysis and calculations were done using Arc/Info 8 and its programming environment. For assessing the positional accuracy of the plots, the following components were investigated:

- positional error of the installed pegs
- distance between neighboring pegs
- plot size (area)
- angle at corners of the square plots

Additionally the percentage of correctly located plot area and the influence of the errors on areal summary statistics for the botanical classification used by Leuzinger (1999) were investigated. The classification was used in the context of habitat preferences in red deer distribution. The summary statistics gives an accurate estimate of the reliability of point in polygon tests based on this data. Point in polygon tests are a common method used in wildlife studies using GIS to determine biotic and abiotic factors at observation locations of animals.

3 Results

Figure 1a shows the distribution of the horizontal offset between reported and true location of the measured pegs. The average error was 5.2 meters (n=194) with a minimum of 0.9m and a maximum of 9.1m. A large part of this error can be attributed to a shift of the set up grid in north direction. After removing this systematic shift the average offset from the true locations remained 2.1m with a minimum of 0.2m and a maximum of 5.6m (Figure 1e).

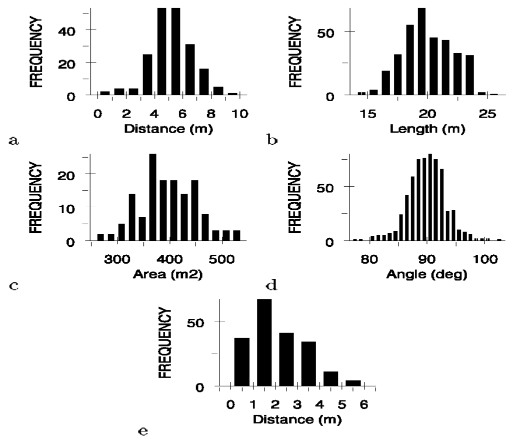

Figure 1 a: Distribution of the horizontal offset between reported and true location of the measured pegs. b: Distribution of (horizontal) distances between neighboring pegs. c: Distribution of the planimetric area of the plots. d: Distribution of the horizontal angle in the corners of the plots. e: Distribution of horizontal offset between reported and true location of the measured pegs after adjustment for the average shift.

The reported (and planned) distance between neighboring pegs is 20 meters. The average (horizontal) distance of the true locations (n=335) is exactly the same (20.0m) with a standard deviation of 2.1m. The minimum distance is 14.8m and the maximum distance is 25.4m. Only about 10% of this distorsion are attributable to altitudinal difference between two corresponding pegs in this relatively flat alpine meadow. The distribution is shown in Figure 1b. The ratio of maximum to minimum length is 1:1.72.

The size of the individual plots was planned and reported to be 400m^2. The average planimetric area of the measured 141 plots was 396.8m^2 with a standard deviation of 54.5m^2 (Figure 1c). The minimum size was 258.6m^2, the maximum size 535.1m^2. The ratio of maximum to minimum size is 1:2.07. The mean horizontal angle in the corners of the plots was 89.9° (n=615) with a standard deviation of 3.3° (Figure 1d). The minimum angle was 77.4° and the maximum was 102.3°.

To assess the correspondance of planned (and reported) versus true positions of plots, we calculated the proportion of a reported plot that was correctly georeferenced (Figure 2a). In an overall statistics, 67.4 percent of the area of all measured plots references the correct area in the field. The average part of the plots that was correctly georeferenced was 269.5m^2 of the total 400m^2 (standard deviation 32.3m^2). The minimal amount was 203.6m^2, the maximum amount 367.2m^2 (Figure 2b).

These values give a good impression about the quality of point in polygon tests. In a random distribution, about 67.4 percent of locations would be correctly assigned to the correct corresponding plots.

In the last calculations made here we test the effects of the above errors on area calculations. We therefore calculated the total area used by each vegetation class according to Leuzinger (1999). She assigned a vegetation class to each of the 268 plots. We calculated the correct area for each class in the 141 plots surveyed here and compared them to the values resulting from the reported sizes (400m^2 per plot). The results are shown in Table 1.

The percentage of area covered by each of the four vegetation classes are almost identical for the planning data and the true measurements. The maximum difference was 0.58 percent for class 3.

Table 1 Cumulate areas (reported and measured) for the 4 vegetation classes used by Leuzinger 1999.

Vegetation Class	Plots (n)	Reported Area (m^2)	Measured Area (m^2)	Reported Area (%)	Measured Area (%)	Difference (%)
1	13	5200	4941.42	9.22	8.83	-0.39
2	38	15200	14926.96	26.95	26.68	0.27
3	49	19600	19765.37	34.75	35.33	0.58
4	41	16400	16311.55	29.08	29.16	0.08
Total		56400	55945.33			

4 Discussion

The results show that for the original studies themselves (Achermann, 2000; Leuzinger, 1999), the area estimations of vegetation units based on the 20m x 20m grid for the whole study area are very accurate. This was expected as this measure is mainly depending on the number of sampling plots and not so very much on the accurate location of each single plot. A higher number of samples naturally results in a better area estimation whereas a higher number of (vegetation) classes results in a decreased areal accuracy. The amount of positional error in the individual pegs measured is astonishingly high. The average error of 5.2m is to a large part explained by local systematic shifts that occurred in the set up of the grid layout. This is a known problem occurring with this kind of set up, as the individual pegs are not positioned independently. Nonetheless the general grid setup was maintained in such a way that the true plots were overlapping at least 50 percent with the planned setup. This was presumably achieved by using an orthophoto to correct for major distorsions.

As stated above overall statistics about vegetation composition for the whole Alp Stabelchod can be made without much error. The main problem with such large positional errors is that the individual sampling plots cannot be determined with reasonable precision. A position 5 or more meters away from the original location may well belong to a different vegetation class. This makes it impossible to revisit the same plot and assess possible changes in the vegetation over time. Changes cannot be distinguished from differences which are due to sampling 5 meters away from the original plot.

5 Conclusions and Recommendations

Considering that for the research in the SNP it is important to enable comparisons over long periods of time, the positional accuracy of research plot documentation should be ameliorated. By determining the exact locations of plots and temporarily used pegs, poles or traps which are dug into the ground, one can also prevent other researchers from obtaining wrong results. Such installations can alter different aspects in the first few centimeters or decimeters of the soil. These locations can be avoided for soil sampling or similar techniques when the exact locations of previous sampling procedures are exactly known.

Point in polygon procedures as have been used by Achermann (2000) for deer locations within individual plots should be used with caution. Considering the variability in the size of the individual plots, it may well be indicated to perform sensitivity analyses (Stoms et al, 1990; Heuvelink, 1998) when calculating animal densities for each individual plot with the original data. The validity of original studies is probably not influenced very much by the amount of positional error found in this accuracy study. By contrast, to achieve true ability to fully reproduce the results and for monitoring changes over long time spans, higher positional accuracies are needed.

In the interest of longterm research the locations of installations and sampling plots should be determined as accurate as possible and feasible. The amount of additional work needed (two days in this case) is small compared to the total duration of a study (four years). For the large scale research areas such as the Alp Stabelchod we recommend that an accuracy of a few centimeters should be achieved. The best and most efficient way is probably to use surveying techniques to determine these locations.

Acknowledgements

We would like to thank Gerald Achermann and Eliane Leuzinger for the permission to use their data for this study. Warmest thanks go to Omar Stornetta, Manuela Bissig, and Peter Egger who kindly participated in the field work. The Swiss National Park authorities gave permission to carry out the field work in this exquisite wilderness area.

References

Achermann, G., 2000, The influence of red deer (Cervus elaphus L.) upon a subalpine grassland ecosystem in the Swiss National Park. PhD thesis, Swiss Federal Institute of Technology, Zurich.

Allgöwer, B. and Bitter, P., 1992, Konzeptstudie zum Aufbau eines Geographischen Informationssystemes für den Schweizerischen Nationalpark (GIS-SNP). Technical report, Wissenschaftliche Nationalparkkommission, Nationalparkdirektion.

Besson, A., 1999, Valorisation des données historique des Rhopalocères du Parc National Suisse. Master's thesis, Université Lausanne.

Bigler, C., 1997, Auswirkungen von Huftieren auf die Wiederbewaldungsdynamik auf drei subalpinen Weiden im Schweizerischen Nationalpark. Master's thesis, ETH Zürich.

Braun-Blanquet, J., 1931, Vegetationsentwicklung im Schweizerischen Nationalpark. Ergebnisse der Untersuchung von Dauerflächen. In *Jahresbericht Naturforschende Gesellschaft Graubündens* 69, pp. 3-82. Naturforschende Gesellschaft Graubündens.

Goodchild, M. and Hunter, G., 1997, A simple positional accuracy measure for linear features. *International Journal of Geographical Information Systems*, 11(3), pp. 299–306.

Goodchild, M. F.,1995, Attribute accuracy. In *Elements of spatial data quality*, Guptill and Morrison (Eds.), pp. 59-79, Elsevier Science, Amsterdam.

Goodchild, M. F. and Gopal, S. (Eds.), 1989, *The Accuracy of Spatial Databases*. London: Taylor and Francis.

Heuvelink, G., 1998, *Error Propagation in Environmental Modelling*. London: Taylor and Francis.

Holzgang, O., 1997, Herbivore-carrying capacity of grasslands in the Swiss National Park. PhD thesis, Swiss Federal Institute of Technology, Zurich.

Hunter, G. and Goodchild, M., 1995, A new model for handling vector data uncertainty in geographic information systems. In *URISA* 95 (1), pp. 410-419, San Antonio, TX.

Krüsi, B., Schütz, M., Wildi, O., and Grämiger, H., 1995, Huftiere und botanische Vielfalt im Schweizerischen Nationalpark. *Inf.bl. Forsch.bereiches Landsch. ökol*. WSL.

Leuzinger, E., 1999, Nachts auf Stabelchod. Das zeitlich-räumliche Nutzungsmuster der Rothirsche (Cervus elaphus L.) auf der subalpinen Weide Stabelchod im Schweizerischen Nationalpark. Master's thesis, Universität Zürich.

Lexer, M.J., and Hönninger, K, 2004, Effects of error in model input: experiments with a forest patch model. *Ecological Modelling* 173, pp. 159-176.

Maggini, R., 1999, Etude de la distribution de Formica exsecta Nyl. à l'aide d'un Système d'information géographique (SIG). *Cratschla* 2/99, p. 22.

Rettie, W.J., and McLoughlin, P.D., 1999, Overcoming radiotelemetry bias in habitat-selection studies. *Canadian Journal of Zoology* 77, pp. 1175-1184.

Salgé, F., 1995, Semantic accuracy. In *Elements of spatial data quality,* Guptill and Morrison (Eds.), pp. 139-151, Elsevier Science, Amsterdam.

Sbresny, J., 1997, Fehlerquellen in Raumbezogenen Informationssystemen. In *Geologisches Jahrbuch, Reihe F, Bodenkunde*. Bundesanstalt für Geowissenschaften und Rohstoffe, Staatliche Geologische Dienste in der Bundesrepuplik Deutschland, Hannover.

Schütz, M., Risch, A.C., Leuzinger, E., Küsi, B.O., Achermann, G, 2003, Impact of herbivory by red deer (Cervus elaphus L.) on patterns and processes in subalpine grasslands in the Swiss National Park. *For. Ecol. Manage*. 181, pp. 177-188.

Schütz, M., Wildi, O., Krüsi, B., Achermann, G., and Grämiger, H., 1998, Zeitreihenanalyse in der Vegetationskunde: Analyse und Interpretation von Einzelflächen am Beispiel von Daten aus dem Schweizerischen Nationalpark. *Bot. Helv*. 108/1.

Skidmore, A. and Turner, B., 1992, Map accuracy assessment using line sampling. *Photogrammetric Engineering and Remote Sensing*, 58(10), pp. 1453-1457.

Stoms, D., Davis, F.W., and Cogan, C., 1990, Sensitivity of habitat models to uncertainties in gis data: a california condor case study. In *Proceedings GIS/LIS* 90 (1), pp. 69-78, Anaheim CA.

WNPK, 1916, Reglement der Kommission für die wissenschaftliche Erforschung des Nationaparkes (W.N.P.K.) der Schweizerischen naturforschenden Gesellschaft.